

[ABOUT](#)
[BROWSE](#)
[FAVORITES](#)
[ALERT](#)
[ORDERS](#)

> Home / Publication / Volume /

Chapter



Lecture Notes in Computer Science

Publisher: Springer-Verlag Heidelberg

ISSN: 0302-9743

Subject: Computer Science

Volume 1280 / 1997

Title: Advances in Intelligent Data Analysis. Reasoning about Data: Second International Symposium, IDA-97, London, UK, August 1997. Proceedings

Editors: X. Liu, P. Cohen, M. Berthold (Eds.):

Chapter: p. 17

Online Date: May 2003

[Previous article](#)
[Next article](#)

Linking Options

Full Text Available

The full text of this article is available. You may view the article as a PDF.

The size of this document is 100 kilobytes. Although it requires a longer download, the PDF is the authoritative online version.

[Open](#)

Decomposition of Heterogeneous Classification Problems

12800017^{A1}, 12800017^{A1}, Se June Hong^{A1}, Jonathan R. M. Hosking^{A1}, Jorge Lepre^{A1}, Edwin P. D. Pednault^{A1}, Barry K. Rosen^{A1}

^{A1} IBM Research Division, T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, U.S.A.

Abstract:

In some classification problems the feature space is heterogeneous in that the best features on which to base the classification are different in different parts of the feature space. In some other problems the classes can be divided into subsets such that distinguishing one subset of classes from another and classifying examples within such subsets require very different decision rules, involving different sets of features. In such heterogeneous problems, many modeling techniques (including decision trees, rules, and neural networks) evaluate the performance of alternative decision rules by averaging over the entire problem space, and are prone to generating a model that is suboptimal in any of the regions or subproblems. Better overall models can be obtained by splitting the problem appropriately and modeling each subproblem separately.

This paper presents a new measure to determine the degree of dissimilarity between the decision surfaces of two given problems, and suggests a way to search for a strategic splitting of the feature space that identifies regions with different characteristics. We illustrate the concept using a multiplexor problem.

For assistance inside the Americas: springerlink@springer-ny.com, For assistance outside the Americas: springerlink@springer-verlag.de
 Springer-Verlag Heidelberg | Tiergartenstr. 17 | D-69121 Heidelberg | Germany | [Privacy](#), [Disclaimer](#), [Term](#)
 Copyright Information

Remote Address: 151.207.244.3 • Server: MPWEB05

HTTP User Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 4.0; T312461)

Update rules for parameter estimation in Bayesian networks (1997) (Make Corrections) (18 citations)

Eric Bauer, Daphne Koller, Yoram Singer

Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI)

View or download:

att.com/~singer/papers/bayesn.ps.gz

stanford.edu/~koller/paper...uai97em.ps

cs.huji.ac.il/~singer/pa...bayesn.ps.gz

Cached: [PS.gz](#) [PS](#) [PDF](#) [Image](#) [Update](#) [Help](#)

CiteSeer
Electronic Literature Digital Library

[Home/Search](#) [Bookmark](#) [Context](#) [Related](#)

From: att.com/~singer/pub (more)

From: pitt.edu/~dsl/UAI97 Bauer UAI97

(Enter author homepages)

(Enter summary)

Rate this article: 1 2 3 4 5 (best)

[Comment on this article](#)

Abstract: This paper re-examines the problem of parameter estimation in Bayesian networks with missing values and hidden variables from the perspective of recent work in on-line learning [12]. We provide a unified framework for parameter estimation that encompasses both on-line learning, where the model is continuously adapted to new data cases as they arrive, and the more traditional batch learning, where a pre-accumulated set of samples is used in a one-time model selection process. In the batch case,... ([Update](#))

Context of citations to this paper: [More](#)

.... execute only a partial maximization step of the EM algorithm based on gradient techniques leading to generalized EM algorithms [8, 10, 11, 9]. In this paper, we report on an experimental comparison between EM and (accelerated) conjugate gradients to maximum likelihood...

.... execute only a partial maximization step of the EM algorithm based on gradient techniques leading to generalized EM algorithms [Thi95,MK97,BKS97,OK99] Finally, KL02] gave experimental evidence that advanced gradient methods namely scaled conjugate gradients are comparable to...

Cited by: [More](#)

Scaled CGEM: A Fast Accelerated EM - Fischer, Kersting ([Correct](#))

Adaptive Overrelaxed Bound Optimization Methods - Salakhutdinov, Roweis ([Correct](#))

Collective Mining of Bayesian Networks from Distributed.. - Chen, Sivakumar.. ([Correct](#))

Active bibliography (related documents): [More](#) [All](#)

0.3: A New Parameter Estimation Method for Gaussian Mixtures - Singer, Warmuth (1998) ([Correct](#))

0.2: A Comparison of New and Old Algorithms for A Mixture.. - Helmbold, Schapire.. (1997) ([Correct](#))

0.2: On the Convergence of Stochastic Iterative Dynamic.. - Jaakkola, Jordan (1994) ([Correct](#))

Similar documents based on text: [More](#) [All](#)

0.2: BoostTexter: A Boosting-based System for Text Categorization - Schapire, Singer (2000) ([Correct](#))

0.2: Reducing Multiclass to Binary: A Unifying Approach for.. - Allwein, Schapire.. (2000) ([Correct](#))

0.2: Sampling in Factored Dynamic Systems - Koller, Lerner (2000) ([Correct](#))

Related documents from co-citation: [More](#) [All](#)

12: Maximum Likelihood from Incomplete Data via the EM Algorithm (context) - Dempster, Laird et al. - 1977

8: Adaptive probabilistic networks with hidden variables - Binder, Koller et al.

6: Neural Networks for Pattern Recognition (context) - Bishop - 1995

BibTeX entry: ([Update](#))

E. Bauer, D. Koller, and Y. Singer. Update rules for parameter estimation in Bayesian networks. In Proceedings of the 13th Annual Conference on Uncertainty in AI, pages 3--13, 1997.

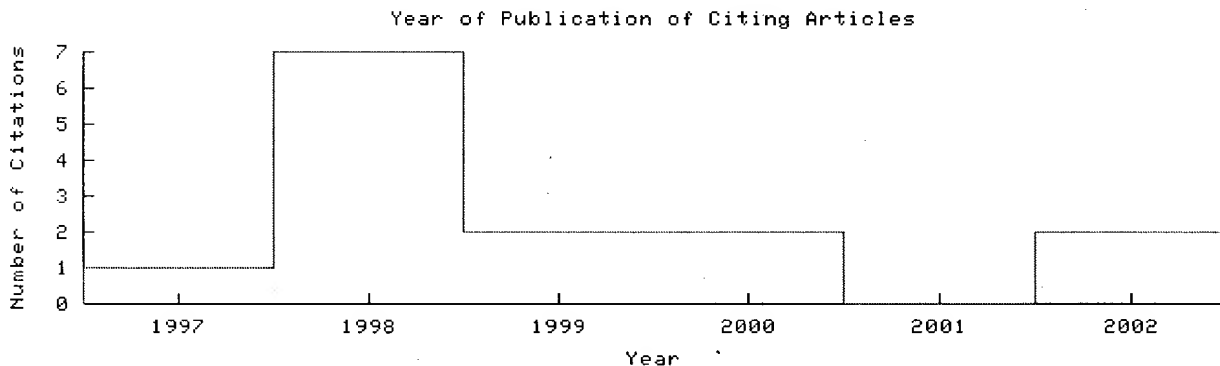
<http://citeseer.ist.psu.edu/bauer97update.html> [More](#)

```
@inproceedings{ bauer97update,
  author = "E. Bauer and D. Koller and Y. Singer",
  title = "Update Rules for parameter estimation in {B}ayesian networks",
  booktitle = "Proceedings of the 13th Annual Conference on Uncertainty in {AI}"
  publisher = "~",
```

```
address = "~",
editor = "~",
year = "1997",
url = "citeseer.ist.psu.edu/bauer97update.html" }
```

Citations (may not include all citations):

- 1735 Maximumlikelihood from incomplete data via the EM algorithm (context) - Dempster, Laird et al. - 1977
- 1612 Elements of Information Theory (context) - Cover, Thomas - 1991
- 511 Cambridge University Press (context) - Horn, Johnson - 1985
- 215 Information Theory and Statistics - Kullback - 1959
- 118 How to use expert advice (context) - Cesa-Bianchi, Freund et al. - 1993
- 103 Operations for learning with graphical models - Buntine - 1994
- 85 The EM algorithm for graphical association models with missi.. (context) - Lauritzen - 1995
- 78 The ALARM monitoring system: A case study with two probabili.. (context) - Beinlich, Suermondt et al. - 1989
- 68 Adaptive probabilistic networks with hidden variables - Binder, Koller et al. - 1997
- 56 Local learning in probabilistic networks with hidden variabl.. - Russell, Binder et al. - 1995
- 40 Inference and missing data (context) - Rubin - 1976
- 20 A comparison of new and old algorithms for a mixture estimat.. - Helmbold, Schapire et al. - 1997
- 19 Accelerated quantification of bayesian networks with incompl.. - Thiesson - 1995
- 18 An iterative procedure for obtaining maximum-likelihood esti.. (context) - Peters, Walker - 1978
- 14 line portfolio selection using multiplicative updates - Helmbold, Schapire et al. - 1996
- 11 On stochastic approximation (context) - Dvoretzky - 1956
- 11 aHUGIN: A system for creating adaptive causal probabilistic .. (context) - Olesen, Lauritzen et al. - 1992
- 10 Eigenvalues of covariance matrices: Application to neural- n.. (context) - Le Cun, Kanter et al. - 1991
- 8 The numerical evaluation of the maximum-likelihood estimates.. (context) - Peters, Walker - 1978
- 8 A stochastic approximation model (context) - Robbins, Monro - 1951



The graph only includes citing articles where the year of publication is known.

Documents on the same site (<http://www.research.att.com/~singer/pub.html>): [More](#)

- On the Learnability and Usage of Acyclic Probabilistic.. - Ron, Singer, Naftali (1995) ([Correct](#))
- Improved Boosting Algorithms Using Confidence-rated Predictions - Schapire, Singer (1999) ([Correct](#))
- Dynamical Encoding of Cursive Handwriting - Singer, Tishby (1994) ([Correct](#))

[Online articles have much greater impact](#) [More about CiteSeer](#) [Add search form to your site](#) [Submit documents](#) [Feedback](#)

CiteSeer - Copyright [NEC](#) and [IST](#)

Unknown Attribute Values In Induction (1989) ([Make Corrections](#)) ([46 citations](#))

J. R. Quinlan Basser Department of Computer Science University of Sydney...
 Cached: [PS.gz](#) [PS](#) [PDF](#) [Image](#) [Update](#) [Help](#)

View or download:

unsw.edu.au/~quinlan/q.ml89.ps[Home/Search](#) [Bookmark](#) [Context](#) [Related](#)

From: unsw.edu.au/~quinlan/ ([more](#))
 (Enter author homepages)

[\(Enter summary\)](#)

Rate this article: 1 2 3 4 5 (best)

[Comment on this article](#)

Abstract: Simple techniques for the development and use of decision tree classifiers assume that all attribute values of all cases are available. Numerous approaches have been proposed with the aim of extending these techniques to cover real-world situations in which unknown attribute values are not uncommon. This paper compares the effectiveness of several approaches as measured by their performance on a collection of datasets. **INTRODUCTION** The 'standard' technique for constructing a decision tree... ([Update](#))

Context of citations to this paper: [More](#)

...rules, Missing Values, Preprocessing, Decision Trees. **1 Introduction** The missing values problem is an old one for analysis tasks[9] [12]. The waste of data which can result from casewise deletion of missing values, obliges to propose alternatives approaches. A current one is to...

...as that of the defining boolean function. **There has been some empirical work studying the task of learning from incomplete data [8, 35, 15].** With the goal of giving a theoretical explanation for the observed empirical phenomena, Schuurmans and Greiner [38, 39] studied the...

Cited by: [More](#)

Classification by Voting Feature Intervals - Demiröz, Güvenir (1997) ([Correct](#))
 Bankruptcy Prediction Using Feature Projection Based.. - Güvenir, Altıngöç (1997) ([Correct](#))
 MRDTL: A multi-relational decision tree learning algorithm - Leiva ([Correct](#))

Active bibliography (related documents): [More](#) [All](#)

0.2: Trends in Topological Combinatorics - Kozlov ([Correct](#))
 0.2: Schröder Paths and Zebra Parallelogram Polyominoes - Sulanke ([Correct](#))
 0.2: Feasibility Study of Radiometry for Airborne Detection of.. - The Nasa Sti (2001) ([Correct](#))

System load high. Please wait...

Timeout. Please try your query later.

Similar documents based on text: [More](#) [All](#)

0.0: MDL and Categorical Theories (Continued) - Quinlan ([Correct](#))
 0.0: FOIL: A Midterm Report - Quinlan, Cameron-Jones (1993) ([Correct](#))
 0.0: Combining Instance-Based and Model-Based Learning - Quinlan (1993) ([Correct](#))

Related documents from co-citation: [More](#) [All](#)

20: Programs for machine learning (context) - Quinlan - 1993
 17: Classification and Regression Trees (context) - Breiman, Friedman et al. - 1984
 17: Induction of Decision Trees (context) - Quinlan - 1986

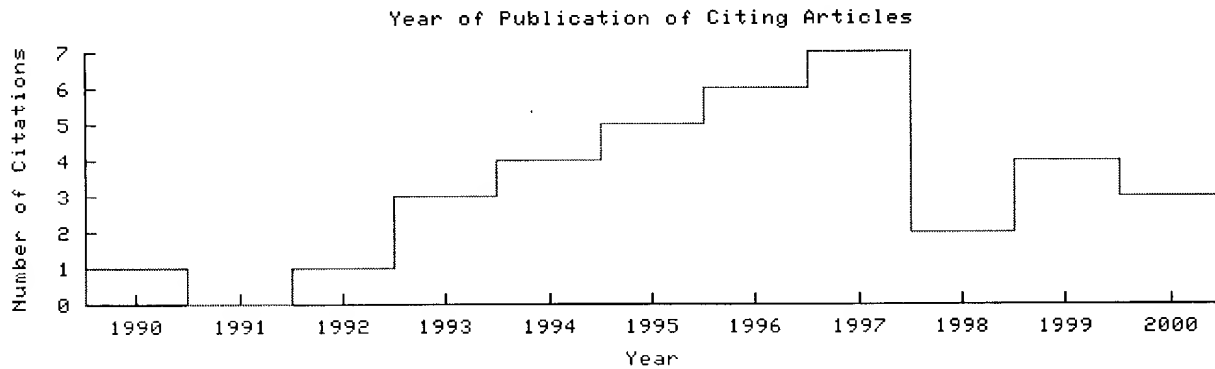
BibTeX entry: ([Update](#))

Quinlan, J. (1989). Unknown attribute values in induction. In Segre, A. (Ed.), Proceedings of the Sixth International Machine Learning Workshop Cornell, New York. Morgan Kaufmann.
<http://citeseer.ist.psu.edu/quinlan89unknown.html> [More](#)

```
@inproceedings{quinlanquinlanunknown,
  author = "J. R. Quinlan",
  title = "Unknown Attribute Values in Induction",
  url = "citeseer.ist.psu.edu/quinlan89unknown.html" }
```

Citations (may not include all citations):

- 1188 Induction of decision trees (context) - Quinlan - 1986
- 1075 Classification and regression trees (context) - Breiman, Friedman et al. - 1984
- 164 The CN2 induction algorithm (context) - Niblett - 1989
- 146 Simplifying decision trees - Quinlan
- 20 Decision trees as probabilistic classifiers (context) - Quinlan
- 16 A recursive partitioning decision rule for nonparametric cla.. (context) - Friedman - 1977
- 5 private communication (context) - Shapiro - 1983
- 1 What do expert systems offer the science of AI (context) - Buchanan - 1989
- 1 Experiments in automatic learning of medical disgnostic rule.. (context) - Kononenko, Bratko - 1984



The graph only includes citing articles where the year of publication is known.

Documents on the same site (<http://www.cse.unsw.edu.au/~quinlan/>): [More](#)

Boosting First-Order Learning - Quinlan (1996) ([Correct](#))

MDL and Categorical Theories (Continued) - Quinlan ([Correct](#))

Oversearching and Layered Search in Empirical Learning - Quinlan, Cameron-Jones (1995) ([Correct](#))

[Online articles have much greater impact](#) [More about CiteSeer](#) [Add search form to your site](#) [Submit documents](#) [Feedback](#)

CiteSeer - Copyright [NEC](#) and [IST](#)